Proposal Tracking Number: 225566

Cross Cut HEP Tracking Algorithms from LHC to LBNF

Josh Bendavid, Pietro Perona, Maria Spiropulu, Jean-Roch Vlimant
California Institute of Technology,
1200 East California Blvd, Pasadena, CA 91125;

Giuseppe Cerati[1], Lindsey Gray, Jim Kowalkowski, Panagiotis Spentzouris
Enrico Fermi National Lab,
PO Box 500, Batavia, IL 60510-5011;

Paolo Calafiura, Steve Farrell, Mayur Mudigonda, Prabhat,
Lawrence Berkeley National Lab,
1 Cyclotron Rd, Berkeley, CA 94720.

Lead PI: Paolo Calafiura, 510-4866717, pcalafiura@lbl.gov
Administrative POC: Isabel Leon, 510-4866353, ileon@lbl.gov

[1] Starting August 2016

Proposal Tracking Number: 225566

Contents

1. Introduction

We are proposing a one-year pilot research program, within the DOE HEP Center for Computational Excellence (HEP-CCE), that will inform the further development of novel advanced pattern recognition algorithms for the LHC tracking detectors with further applications in HEP tracking at large.

Received Date: 6/27/2016

Tracking algorithms measure the curvature of the trajectories of charged particles as they propagate in a magnetic field. From the curvature one can deduce the charge and momentum of the electrons and muons produced, for example, in an ultra-relativistic proton-proton collision at the LHC. This allows separating interesting new physics processes (e.g. production of a supersymmetric particle) from Standard Model background signals. Typically tracking consists of a pattern recognition step, in which the signals left by charged particles as they propagate out from the interaction point through the detector are connected into a track candidate; followed by a track fitting step in which the track candidate trajectories are fitted against a detailed track propagation model. A brief introduction to LHC Tracking can be found in Appendix 6, and there are many extensive reviews of this subject [FRUB, RAGU].

This pilot will provide an end-to-end strategy to optimize LHC tracking algorithms, from detector data simulation, to a reference tracking solution against which tracking algorithms can be evaluated and validated. We will demonstrate the effectiveness of this approach in the pilot project by developing an optimized, scalable track formation algorithm. Additionally the pilot will guide tracking algorithm developments for Liquid Argon Time Projection Chamber based (LArTPC) experiments that will measure precisely the neutrino oscillation parameters and investigate the origin of CP violation

The research team is a partnership between the high energy physics and computer science communities within DOE which includes members of two LHC experiments (ATLAS and CMS), with tracking expertise and access to development efforts from both experiments, members of the LArTPC community, as well as computer scientists with expertise in large-scale data analytics, computational vision, and statistical learning from Berkeley and Caltech.

2. Research Goals and Impact

2.1 Summary of Goals and Objectives

Here, we propose a one-year pilot project to evaluate and broaden the range of computational techniques and algorithms utilized in addressing the LHC upgrade tracking challenge. The pilot will be executed by a partnership between the high-energy physicists and computer scientists at Berkeley, Fermilab and Caltech. Specifically this one-year pilot will provide a framework (Figure 1) to develop and evaluate new algorithms for track finding and classification. The framework will be demonstrated by applying advanced pattern recognition techniques to track candidate formation.

2.2 Impact

We believe that incremental optimization of current LHC tracking algorithms has reached the point of diminishing returns. These algorithms will not be able to cope with the 10-100x increase in HL-LHC data rates anticipated to exceed O(100) GB/s by 2025, without large investments in computing hardware and software development or without severely curtailing the Physics reach of HL-LHC experiments. An optimized track formation algorithm that scales linearly with LHC luminosity, rather than quadratically or worse, may lead by itself to an order of magnitude improvement in the track processing throughput without affecting the track identification performance, hence maintaining the physics performance intact.

The one year pilot-project we propose will enable and inform further research necessary to sustain the physics mission of the LHC experiments through 2030. It will also inform our long-term vision to develop robust, efficient, and scalable full precision tracking algorithms applicable to LHC as well as other current and future experiments -- for example LArTPC-based neutrino experiments, medical imaging, and nuclear physics experiments.

This pilot is well aligned with HEP-CCE mission to foster a “more common HEP computing environment and when possible move away from experiment-specific software”. This proposal is also in alignment with the recommendations produced by the ASCAC Subcommittee on Synergistic Challenges in Data-Intensive Science and Exascale Computing. Specifically, the ASCAC report calls for data analytics integration with exascale computation systems, towards new kinds of workflows that will impact both data-intensive science and exascale computing.

This pilot will benefit the ASCR mission by demonstrating novel data-intensive workflows and proto-validation and verification of those. As we will discuss in detail in section 3.3, we will explore the applicability of **advanced machine learning algorithms beyond domains such as linguistics and neurosciences, into precision tracking applications in the presence of high data volumes and combinatorial ambiguities**.

3. Research Plan

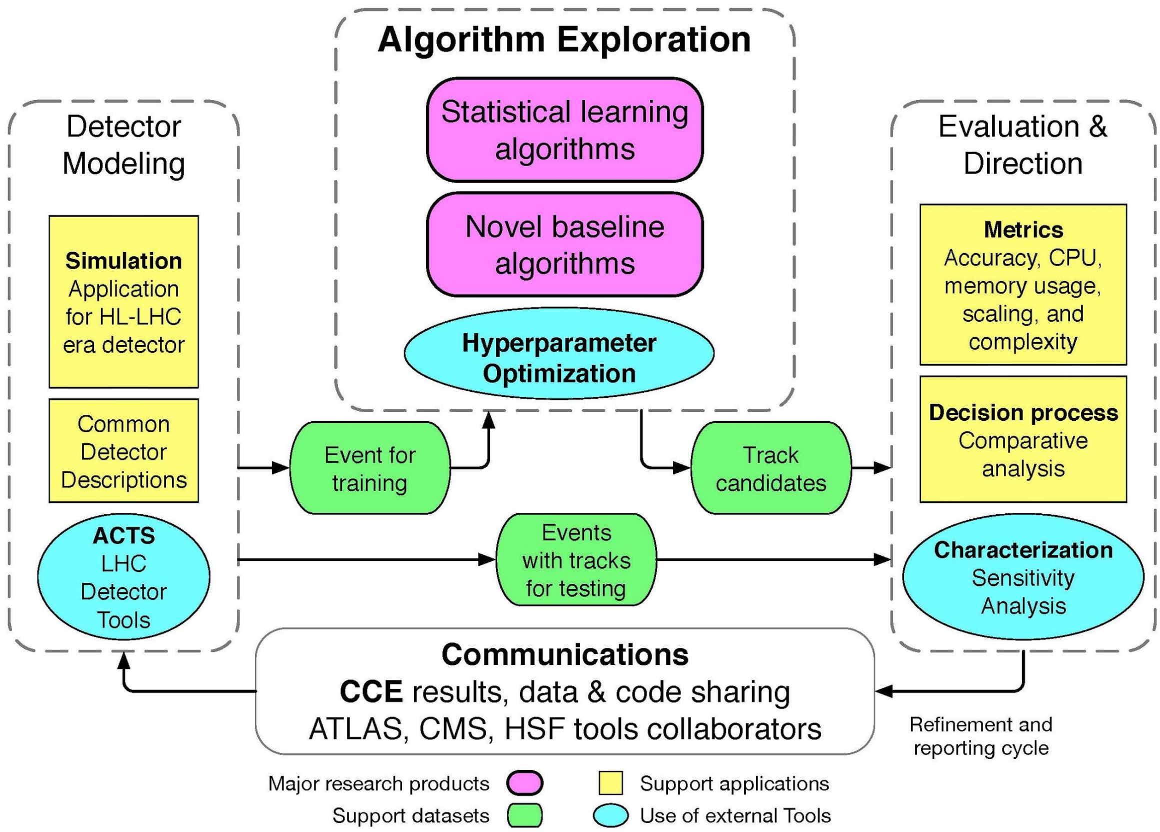


Figure 1: High-level structure and components for this research project

Our strategy to investigate pattern recognition for HL-LHC tracking detectors high-level structure and components for this research project is outlined diagrammatically in Figure 1. This pilot project will test the envisioned design in Figure 1 **using the most computationally complex and resource intensive step of charged particle tracking, namely the track candidate formation step**. The *Detector Modeling* task provides us with simulated event data representative of what is expected from the tracking detectors in the HL-LHC era. We will use a generic collider geometry compatible with an HL-LHC detector design noted in the flowchart of Figure 1 as *Common Detector Descriptions.* We will also integrate existing tracking and other physics tools developed by ATLAS and CMS noted as *LHC Detector Tools*, hence leveraging and moving fast towards the implementation of our novel core developments shown at the center of the graph. Since Learning

Algorithms will be a fundamental part of this pilot, our *Simulation Application* must be capable of generating samples both for testing and training. There are two core components within the Algorithm and Methods project tasks*: Statistical Algorithms and Novel Baseline Algorithms.* They are the central focus of this pilot project. Our group will greatly benefit from access to DOE HPC resources to optimize the performance of our algorithms through *Hyperparameter Optimization* techniques that in recent years have been shown to impact significantly the performance of statistical learning algorithms.

The success of the pilot project is quantified within the task set noted as *Evaluation and Direction* in the flowchart of Figure 1.

Precision tracking is a critical and important task within an LHC experiment. It is usually performed after data is written to storage using the production batch system, with accuracy control parameters set to match available computing cycles given a data analysis schedule. Frequent *External Communications* are necessary with the precision tracking experts within the experiments to assure that results can be interpreted and further improvements can be achieved. We will take advantage of the accumulated offline tracking workflow knowledge and experience to implement a competitive, fast, real-time tracking workflow and rank its performance (labeled *Decision Process* task in the graph) after comparative analysis based on *Metrics* developed in the corresponding task.

3.1 Preparing Data Samples

To develop and test the pilot LHC track formation algorithm, a detector geometry description and simulated space-point datasets must be prepared. The detector geometry we will develop has to be generic yet sufficiently sophisticated to be representative of any LHC-like detector. Likewise, the simulated data will be representative of HL-LHC physics conditions (event topologies, particle kinematics), accurately modeling the detector response. Finally, we will simulate events with sufficient statistics and store them in a data format appropriate for the development and testing of the pilot algorithm. To address these requirements, we will implement a standalone simulation/data-preparation workflow in collaboration with the HEP Software Foundation Common Tracking Software (CTS) Forum. The CTS Forum is developing an experiment-independent tracking and geometry software library [aCTS] upon which we will build realistic, generic LHC detector models. We will use the models, along with the Geant4 simulation toolkit to produce simulated data samples. **An order of 10 million simulated events containing 10 billion tracks will be needed for this pilot project, totaling 1 TB of data.**

An energetic hard scatter vertex mixed with O(200) "pile-up" proton-proton collisions provides good statistics for accurate determination of the tracking efficiencies for the primary vertex, while also allowing for tests of tracking capabilities in the low p_T regime of the pile-up interactions, and thus presents an accurate and overall picture of the tracking performance.

The simulation data workflow is an essential first piece of this pilot project. We thus aim to have the complete chain in place within the first three months. We expect development then to be iterative during the algorithm studies.

1. **Deliverable**: realistic full simulation samples of generic HL-LHC tracking detector with a format and size that allows efficient algorithm development

3.2 Novel Baseline Algorithms

We will start our investigation by implementing a best-practice solution for seeding and track candidate formation. We will use this reference solution as a benchmark both in terms of physics performance, and computing performance.

The Kalman filter is a set of equations that recursively estimates the state of a (linear) dynamical system at time t ($\mathbf{a}_t$) by optimally combining noisy predictions based on the system previous state ($\mathbf{a}_{t-1}$) with some number of noisy observations ($\mathbf{b}_t$) (Fig 2).

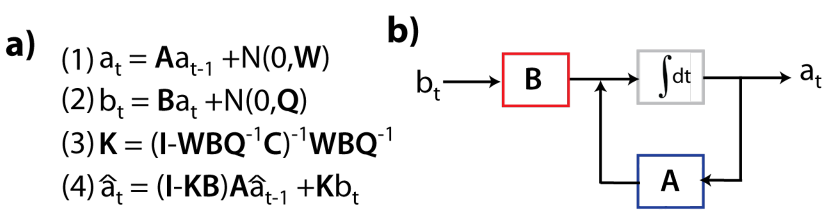


Figure 2: *A Kalman filter operates recursively on streams of noisy input data to produce a statistically optimal estimate of the underlying system state.* ***a)*** *The equations of the Kalman Filter: (1) a 1st order linear dynamical process, in which the state of the system (*$\mathbf{a}_t$*) evolves according to the dynamics matrix* ***A*** *with additive Gaussian noise (model covariance* $\mathbf{W}_t$*). (2) Observations (*$\mathbf{b}_t$*) of the state variable (*$\mathbf{a}_t$*) are assumed to be a linear transform (****B****) of the state, perturbed by additive Gaussian noise (observation covariance* $\mathbf{Q}_t$*). (3) The Kalman Gain (*$\mathbf{K}_t$*) is calculated according to the ratios of the processes (*$\mathbf{W}_t$*) and observation (*$\mathbf{Q}_t$*) noise covariances. (4) The optimal estimate of the system state (*$\mathbf{a}\hat{}_t$*) is produced by combining the estimate from the previous state with the observations (*$\mathbf{b}_t$*), weighted by the Kalman Gain.* ***b)*** *Block diagram of system dynamics and signal combinations in the Kalman Filter.*

Over the last three decades, Kalman Filter has become the classic algorithm for track fitting [FRUE] and formation [CKF]. In track formation, the state of the KF $\mathbf{a_t}$ is represented by the position, direction, and curvature (momentum) of a track at a detector surface **t**. The charged particle propagates from surface **t** to surface **t+1** following a *track model* $\mathbf{f_t}$ which is in general non-linear[2], but can be linearized in the neighborhood of the detector surface **t**. Under this assumption, the first term of $\mathbf{f_t}$ Taylor expansion becomes the KF dynamics matrix $\mathbf{A_t}$.

The "Parallel Kalman Filter" [PKF] developed by a member of our team represents the state of the art in track formation algorithms. Since it is designed to be vectorized by the compiler, and to be run in parallel on multiple threads on many-core platforms, it serves as a good representation of future performance and therefore is the algorithmic benchmark to which we should compare our results. A parallel implementation of the Kalman Filter algorithm optimized to run efficiently on GP-GPUs will be of interest across HEP domains, and to any scientific application that needs to analyze high-volume time series.

2. **Deliverable**: a reference solution for seeding and track candidate formation capturing state of the art algorithms.

3.3 Modeling Track Dynamics

Experimental High Energy Physics has always been a fertile ground for statistical techniques, from the early experimentations with neural networks of the LEP era, to the widespread adoption of Multivariate Analysis [TMVA] for Tevatron and LHC Run1 data analyses, to recent successes in the application of Machine Learning to LHC data analyses culminated into the HiggsML challenge [HIML].

Revolutionary advances in statistical techniques, and the availability of hardware resources like GP-GPUs and SSDs give us the opportunity to approach problems like track seeding or track formation in ways that only five years ago would have appeared to be completely out of reach.

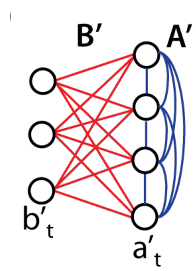
Based on the experience with many physics analysis applications (most notably the discovery of single top at the Tevatron and the Higgs at the LHC), we expect that letting an algorithm classify a seed, or group a list of space-points together in a track

---

[2] The track model represents the Lorentz propagation of a relativistic charged particle of momentum **p** in an inhomogeneous magnetic field **B(x)**. The effects of the particle interaction with detector material are usually represented by an average energy loss, added to the track model, and by a stochastic perturbation of the trajectory coming from multiple scattering which contributes to the model covariance $\mathbf{W_t}$.

candidate, will lead to improved physics performance with respect to a hand-tuned traditional algorithm. Even if these improvements in efficiency or purity were to be incremental, we believe neural network algorithms will offer several advantages over traditional iterative tracking algorithms. Firstly, neural network algorithms are computationally very regular, and lend themselves to run efficiently on data parallel architectures, particularly on SIMD architectures like the vector units of modern processors, and GP-GPU coprocessors. Secondly, neural network algorithms are generally more robust than traditional algorithms when running with reduced precision on low power approximate computing platforms. Finally, statistical learning algorithms can be used to discover new features in data sets they are trained on [JETS].

The problems of track building and fitting are inherently iterative due to the sequential nature of space-point data. A charged particle's propagation through an LHC detector is a sequence of steps determined by the track kinematics as well as randomization due to material scattering effects. Traditionally, the Kalman Filter algorithm is used to model this evolution of the particle state, combining information from the evolving track model with space-point data at each successive detector layer. Recurrent neural networks (RNNs) have enjoyed considerable success at modeling sequence data in various types of applications. RNNs use recurrent connections to carry hidden state or "memory" through successive operations, allowing modeling the dynamic evolution of a state. These algorithms are being heavily used in the areas of language translation [SSL] and speech recognition [FASR, SRDR]. The Kalman Filter algorithm itself has a natural implementation as a recurrent neural network. Specifically, each state **a'$_t$** of the system can be associated with a neuron in a recurrently connected network, and the system dynamics (or, equivalently, the probability of transitioning between states) are determined by the synaptic strengths between neurons (**A'**). The observations **b'$_t$** are feedforward inputs to the network with synaptic weights represented by the observation matrix **B'**.



Several studies from the Brain-Machine Interface community [SUSS] have shown that RNNs can outperform Kalman Filters when modeling the dynamics of nonlinear systems. We will investigate the use of RNNs for building and fitting LHC tracks and we will leverage our experience in this area.

RNNs while theoretically known to be powerful, historically suffered from the problem of *vanishing gradients*, thus making them hard to train. Further, it was challenging with regular RNNs to model complex long-term dependencies in the inputs. The Long Short Term Memory Recurrent Networks [LSTM] address the problem of long-term dependencies in inputs by introducing the idea of *gates*. Commonly, there are three types of gates -- input, forget and output gates. Briefly, what these gates do is to allow for the network to selectively accept inputs or forget inputs. When used in an ensemble, these classes of models can capture the underlying dynamics of many previously unsolved time varying pattern recognition problems.

Given, x_t as the time varying input to the system;

W_i, W_f, W_c, W_o, U_i, U_f, U_c, U_o, V_o are weight matrices (also known as linear operators that transform inputs);

b_i, b_f, b_c and b_o are bias vectors (that capture the first order statistics);

the following are the update equations for the memory cells at each time step t. For the *input gate,* the updates are as follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

$$\hat{C} = tanh(W_c x_t + U_c h_{t-1} + b_c)$$

The *forget gate* is updated thusly, where we weight the candidate inputs from the previous step and f_t, the input response of the forget gate

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

$$C = i_t * \hat{C} + f_t * C_{t-1}$$

Lastly, with the new state of the memory cell we can compute the value of their output gates and subsequently, their outputs:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$h_t = o_t * tanh(C_t)$$

The ability to learn what to *store* and *forget* to reproduce patterns seen in the inputs, gives LSTMs a greater range of expressive power than traditional methods like *Kalman Filters* to discover and model dynamics in the input data. Specifically, to model track patterns, we will use simulated data where we know the trajectories of the particles (in 3D space) and feed them into an LSTM network to train. A trained network could then be seeded with a triplet of 3D space-points of a single particle and used to build a list of space-points (track candidate) compatible with the seed.

HEP tracking input data are unusual in that they consist of a sparse series of high-precision space-point measurements[3]. Typical applications of LSTM networks, such as speech recognition, usually rely on a long, continuous time-series of relatively low precision measurements. Another characteristic of HEP track finding is the availability of a detailed track propagation model, essential to constrain the extrapolation from one detector layer to the next. A challenge for this study will be to find efficient techniques to teach this strong prior knowledge to the LSTM network while maintaining its power to discover autonomously aspects of the track model from the high-precision inputs. Another challenge in applying LSTMs to this problem will be to explore and optimize the architecture of the network. To do this, we plan to use hyper-parameter search methods such as *Spearmint* [SPEA] relying on DOE HPC platforms like NERSC cori to perform the search in a reasonable time.

If successful, this investigation will greatly expand the applicability of RNNs beyond traditional domains such as linguistics and neurosciences, to scientific applications, including HEP tracking, that require precision pattern recognition in the presence of high-volume input data and combinatorial ambiguities.

3. **Deliverable**: Novel pattern recognition algorithm to build track candidates from seeds and space-points.

3.4 Implementation to LArTPC applications

The international long and short baseline neutrino program, which includes MicroBooNE, ICARUS, SBND, and DUNE can also benefit from this pilot project. The LArTPC community faces many computational challenges with 3D tracking and particle identification in the presence of noise and, in some cases (when located on the surface as opposed to underground) the unwanted background of cosmic ray tracks. The results of this pilot in the areas of parallel tracking through Kalman filter advances and Deep Learning will also benefit this community. We will implement the advances we will make utilizing the HL-LHC like infrastructure in the protoDUNE (a prototype detector for DUNE) framework, and utilize protoDUNE Simulations to evaluate performance.

4 **Deliverable**: Adapt track finding algorithm and evaluate in LArTPC environment.

---

[3] A state of the art tracking detector will measure the position of O(10) 3D space-points per track with $O(10^{-5})$ precision.

3.5 Metrics, Analysis

The applicability of this algorithm work to future tracking scenarios can be quantified by computing efficiency and physics performance, including any relative trade off between the two. Typical physics performance metrics for tracking, such as track-finding efficiency and the rate of finding "fake" tracks arising from combinatorics or bad seeding, can be used to determine the baseline viability of the methods being studied. Efficiencies below 90% or fake rates of more than 5%, integrated over the fiducial volume, are not suitable for online selection or analysis. For algorithmic computational efficiency, a resulting list of tracks and hits must be attainable within 300 ms. This time constraint is due to further required steps of sorting and selection of tracks, in order to remove duplicates. It will be possible to tune the "working point", in terms of efficiency vs. fake rate, to produce tracks satisfying the physics performance criteria and the time budget.

Another metric that quantifies stability of an algorithm in realistic detector environments, is the ability to maintain performance when presented with inputs that have been systematically altered with respect to the datasets used for testing and training. This enables quantification of the stability of the algorithms in realistic data-taking scenarios. This "input drift" can be caused by numerous sources, including, alignment of the tracking system, and effects of radiation damage on signals in the sensors of the tracker. The effects of all of these sources of drift can be quantified by their effect on tracking efficiency and fake rate. Computing time is also a useful metric to assess the impact of input drift.

One of the main issues with statistical learning techniques is estimating the systematic error associated with model training, or measuring variation in generated models across many training runs, given the same starting conditions. Training questions that will be addressed here include the sensitivity of the model to (1) variations in test data, (2) changes in training samples class ratios, and (3) changes in the model hyperparameters.

For this pilot project overall, success will be primarily measured using efficiency and physics performance. We will provide initial measurements of input sensitivity and systematic uncertainties.

3.6 Connection to other projects

We will collaborate with the Common Tracking Software (CTS) forum of the HEP Software Foundation. We will use their existing toolkit [aCTS] which includes geometrical description and track extrapolation tools, to setup our shared HL-LHC detector simulation, and as building blocks to implement our reference seeding and track formation chain. We will contribute to CTS improvements to their tools, and share new common tracking tools we may develop.

3.7 Outlook

This pilot will inform our vision of developing advanced pattern recognition algorithms to address the expected resource deficit for charged particle tracking at the LHC and other HEP future experiments including at LBFN. During the pilot we will attack the most pressing issue, which is the track candidate formation algorithm. This is embedded in our vision that encompasses the full chain of charged particle processing:

- filtering in real-time most background events by running precision tracking and vertexing on dedicated hardware in O(1)μs;
- speeding up simulation of charged particle propagation using generative networks instead of traditional Monte Carlo techniques;
- optimizing track seeding algorithms and data structures to address the combinatorial issues deriving from increased LHC luminosity (and energy);
- applying image processing techniques to seed classification, track/space-point association, beam-spot structure detection.

The primary outcome of the pilot will be a first iteration on the framework shown in Figure 1, to develop and evaluate tracking algorithms efficiently, reliably, and in an experiment-independent way with wide applications in HEP. This will enable collaboration with tracking experts from the LHC community and beyond.

Proposal Tracking Number: 225566

4. Project Organization: Coordination, Management and Deliverables

4.1 Management and Coordination

This project will be executed by three groups at CalTech, FermiLab, and LBNL/UC Berkeley. US ATLAS, US CMS will be external partners providing mainly access to data and common software. We will collaborate with other LHC and HEP communities through DOE's HEP Center for Computational Excellence and the HEP Software Foundation.

Given the tight timescale and the nature of our research, this pilot project will need quick development cycles, with continuous integration of our work products, and, crucially, continuous communication among team members. We will use an instance of the Basecamp web-based project-management tool to coordinate our schedules, communicate through chat rooms and mailing lists, as well as to keep records of the deliverables progress, of all minutes, etc. We will maintain our code on HEP-CCE github, and will peer review our software artifacts before pushing them to the main repository. We will have focused technical bi-weekly meetings to prioritize the work for the week and to review open issues. Once a month we will have a longer "strategical" meeting led by the three co-PIs.

To ensure a fruitful collaboration, members of the team will participate regularly to CCE and HSF meetings.

The Caltech CMS group has leading roles and responsibilities in the CMS experiment in the areas of physics, trigger, software, computing & networks as well as future detector R&D. A crucial aspect of the Caltech approach is taming the information flow from the LHC collisions vertically: from triggering, dataset definition, data quality monitoring, software development, to distributed computing and collaborative systems, all the way to data-driven physics analysis with multiple controls for validation and verification towards rapid discovery and characterization of signals of new physics. The group collaborates on campus with the Center for Data-Driven Discovery (CD3) http://cd3.caltech.edu) and DOLCIT (the center for Decision, Optimization, and Learning at the California Institute of Technology http://dolcit.cms.caltech.edu/index.html) and has traditionally very strong working relations with industry as well as research centers and laboratories around the world on agile intelligent systems that integrate learning and planning. The group has launched and is leading in CMS a powerful precision timing detector R&D program and maintain a leading role in the precise calibration of the electromagnetic lead tungstate calorimeter and the mitigation of noise and operation of the hadron calorimeter including the Phase-I HO readout upgrade to SiPMs. The group has played a pivotal role in the preparation and execution of the physics program in the electron, photon, missing energy and jets physics objects, in the trigger and dataset definitions with focus on the triggers for new physics, and in the SUSY, Exotica and Higgs physics groups -- including the work on the Higgs discovery and characterization in diboson final states . Caltech's Tier 2 is the very first Tier 2 of the LHC grid. It was proposed and prototyped in 1999 and it was commissioned and brought in production in 2001. It provided a proof of concept of the Tier 2 and the LHC Data Model. The Tier 2 at Caltech provides substantial and reliable computational and storage resources to US CMS and CMS, (more than 66.8 K of HS06 computing units and 4.8 petabytes of raw storage). It combines production processing of simulated events, as well as support for US CMS physics analysis, and computing, software systems and network development. It is an active part both of the production and computation R&D efforts in CMS. It is the first Tier 2 to commission the 100 Gbps uplink in 2014 and lead the effort to help all US CMS Tier 2 sites reach such a goal. A testbed is presently set up at the Caltech Tier2 in order to improve the support infrastructure for data federations at CMS. As a first step, we have built systems that produce and ingest network data transfers up to 80 Gbps. As part of this project, work within the Caltech group is ongoing to develop a plugin for CMSSW based on libdavix and better interaction of HTTP-over-Xrootd with the OSG distribution.

The Caltech CMS group has also an associated hybrid physics/engineering team with expertize on: (1) state of the art data transfers over long distances, (2) pervasive real-time monitoring of networks and end-systems, (3) autonomous steering and control of large distributed systems using robust agent-based architectures and real-time monitoring of the hundreds XrootD data servers used by the LHC experiments supported MonALISA system, (4) the development of software driven multilayer dynamic circuits, and autonomous optical patch panels which provide a virtualized interconnection service for megadata centers and cloud computing, (5) software defined networking, (6) integration of the network awareness and control into the mainstream of the experiments' data and workflow management as in Caltech's OliMPS and ANSE projects, and (7) the exploration of Named Data Networking as a possible content-centric future architecture replacing the current Internet, together with ESnet and leading groups in climate science led by Colorado State that have deployed an NDN testbed. The Caltech team's expertise is based on many years of engagement in LHC physics, computing and networking. Beyond our strong ties with many academic partners, we have built a large network of partnerships with the major R&E networks in the North and South America, Europe and Asia, as well as network vendors, computer and storage manufacturers, and computing system integrators.

The CMS Caltech team in based at Caltech, FNAL and CERN.

Fermilab is directly involved in the CMS Phase-II upgrade project for the HL-LHC physics program. In addition, the CMS LHC Physics Center (LPC) is located at Fermilab. Scientists within this center are directly involved in physics analysis, tracking software, detector upgrades, and CMS management. The FNAL core software infrastructure group with the Scientific Computing Division (SCD) is the home of the principal authors of the software reconstruction framework used within CMS, CMSSW. This group adapted CMSSW into art, a software framework now used by the majority of Fermilab's muon and neutrino program experiments. Both CMSSW and art use state-of-the-art C++ and provide interfaces and deployment strategies to most of the production physics algorithms in use by CMS and the Fermilab experiments. They follow a long history of transformative collaborative frameworks and software infrastructure developed for the previous generation major HEP experiments CDF and D0. The latest major upgrade of CMSSW for support of multithreading was designed, coordinated, and development by this group. The group is also responsible for major performance improvements throughout the software stack in use by all Fermilab-based experiments. The FNAL SCD reconstruction group has been directly responsible for CMS tracking software upgrades to evaluate vector processing on Intel Xeon Phi. They have also been responsible for developing particle flow algorithms to be used for the CMS Phase-II high granularity calorimeter (HG-CAL) upgrade. The SCD simulation group and experts in Geant4, and run the SciDAC Geant project. SCD is also managing research projects through Computational HEP to make experiment software, including key infrastructure components usable within current leadership computing facilities.
This mix of direct CMS involvement, collider physics, tracking algorithms, particle simulations, and large-scale software framework expertise ensure success of this pilot project.

The LBNL ATLAS Software group has led the development of ATLAS Core Software since 2001, introducing new software paradigms such as component-based software, data-driven application steering, multi-process and multi-thread concurrency. The group has always been at the forefront of HEP application research of novel computing platforms such as Intel Xeon Phi and IBM TrueNorth. The LBNL NERSC Data and Analytics Services group leads the large and diverse NERSC user community in adopting best-of-breed practices and tools for scientific data management, analysis, and visualization. It has received an ASCR award to further the research in highly scalable Statistics and Machine Learning algorithms (MANTISSA). The UC Berkeley Redwood Center for Theoretical Neuroscience develops mathematical and computational models of the underlying neurobiological mechanisms involved in perception, cognition, learning, and motor function. They collaborate with experimental neuroscience labs in the design of experiments and in the analysis of neural data. The project will profit from this unique combination of skills to address effectively its data engineering, and pattern recognition challenges.

4.2 Summary of Deliverables and Milestones

Table 1: Project Deliverables and Milestones

Group	Deliverable	Due	Institutions
Simulation	1: Realistic full simulation samples of generic HL-LHC tracking detector.	Q4 16	LBL, FNAL
Algorithm	2: Reference solution for seeding and track candidate formation using state of the art algorithms.	Q2 17	FNAL, Caltech
	3: Pattern recognition algorithm to build track candidates from seeds and space-points.	Q3 17	Caltech, LBL
	4: Adapt track finding algorithm and evaluate in LArTPC environment	Q4 17	FNAL, Caltech
Milestone: Demonstrate full simulation and reconstruction chain for generic HL-LHC tracking detector and produce plan for full development and implementation		Q4 17	ALL

5. Budget

This pilot project is applying for funding for **one year** from DOE Advanced Scientific Computing Research and High Energy Physics programs.

Caltech: $200K

Caltech will be funded through a subcontract with FNAL

- Jean-Roch Vlimant, a research scientist with expertise in large scale computation systems and data analytics will **work at 40%** on the development of the algorithm and its deployment in pattern recognition.
- Josh Bendavid, a research scientist with expertise in large scale computation and data analytics will **work at 30%** on the pattern recognition algorithm and on the reference algorithm.
- Stephan Zheng a Computer Science PhD student on AI methods at EAS **will work at 40%** on the pattern recognition algorithm and on the reference algorithm and two HEP Physics PhD students will work **at 30% each** on validation studies of the LHC and DUNE applications of the reference algorithm.
- Pietro Perona, Professor of Electrical Engineering and Computer Science at Caltech's EAS Division will **work unfunded** as a senior advisor for this pilot project.
- Maria Spiropulu, Professor of Physics at Caltech's PMA **will work at 5%** on the pattern recognition and reference algorithm as well as the overall pilot project management, and **5% on HEP-CCE** coordination towards the long term strategy.

FNAL: $275K

- Lindsey Grey, an associate scientist with the Scientific Computing Division, will **work at 25%** on the simulation and the reference algorithm.
- Giuseppe Cerati, an associate scientist with the Scientific Computing Division, will **work at 30%** on the pattern recognition algorithm and on the reference algorithm, including a LArTPC application and evaluation
- Jim Kowalkowski, a computer science researcher with the Scientific Computing Division will **work at 15%**, and to work on simulation input data samples representative of HL-LHC physics. He will also **work at 7%** on the LArTPC implementation and HEP-CCE interface and communication of results.
- Panagiotis Spentzouris, scientist and head of the Scientific Computing Division, will **work unfunded**, on the overall project management and as senior advisor for this pilot project.

LBNL: $275K

- Mayur Mudigonda, a graduate student with the UC Berkeley Redwood Center for Theoretical Neuroscience, will **work full time** on designing the pattern recognition algorithm.
- Steve Farrell, a postdoc with LBNL Physics Division, will **work at 55%** on optimizing the inputs, results, and in general the performance of the pattern recognition algorithm.
- Paolo Calafiura, a scientist with LBNL Computing Research Division will **work at 10%** on simulating input data samples representative of HL-LHC physics, as well as the overall pilot project management. He will also **work at 5%** with HEP-CCE to produce a long-term plan for full development and implementation.
- Prabhat, a scientist with NERSC Data & Analytics group, will **work at 10%**, advising the pilot project on how best to exploit NERSC cori to train and test our algorithms.